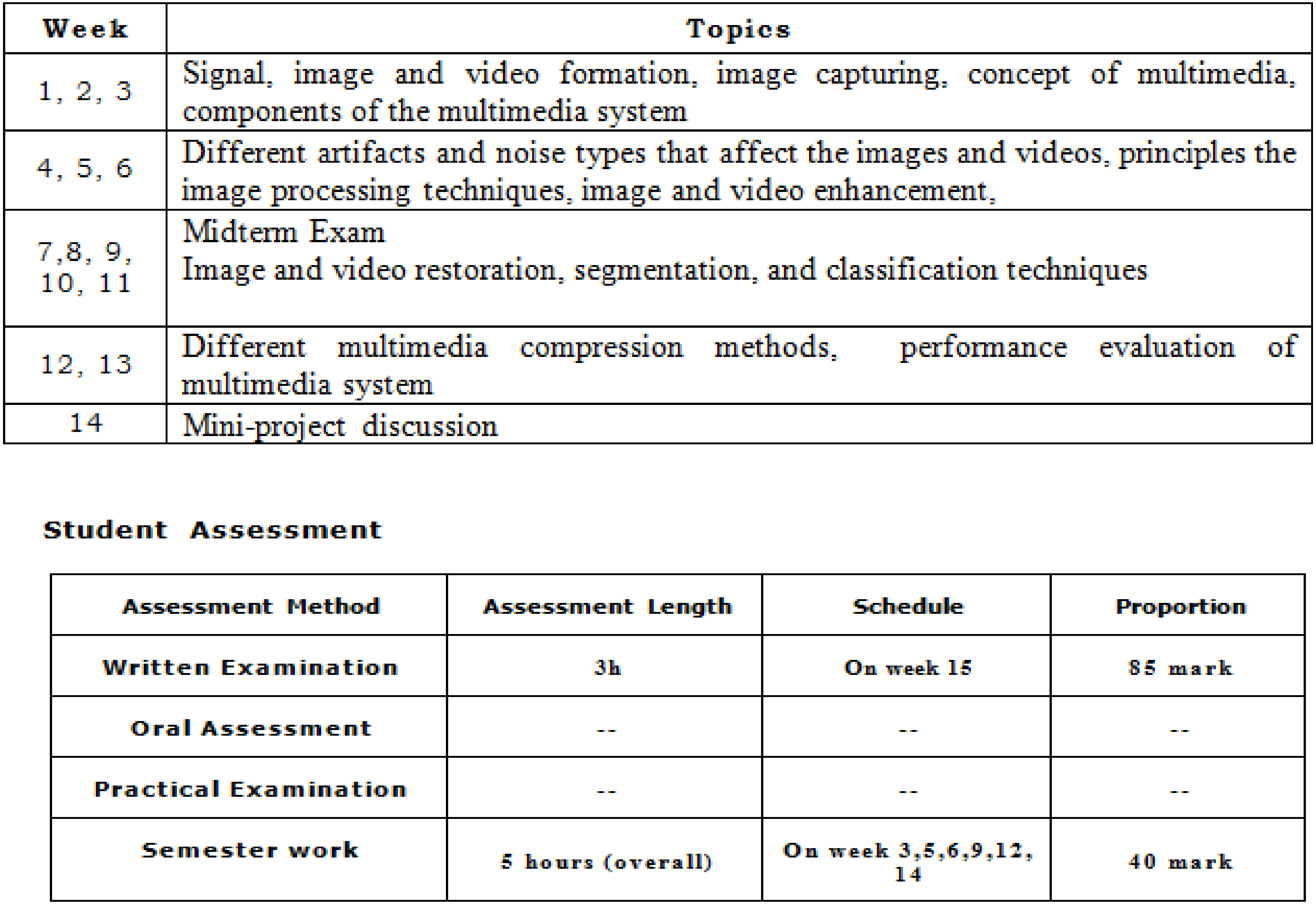
Multimedia and Image Processing

2nd Semester- 2017-18

Lecture 1

Introduction

Course Overview

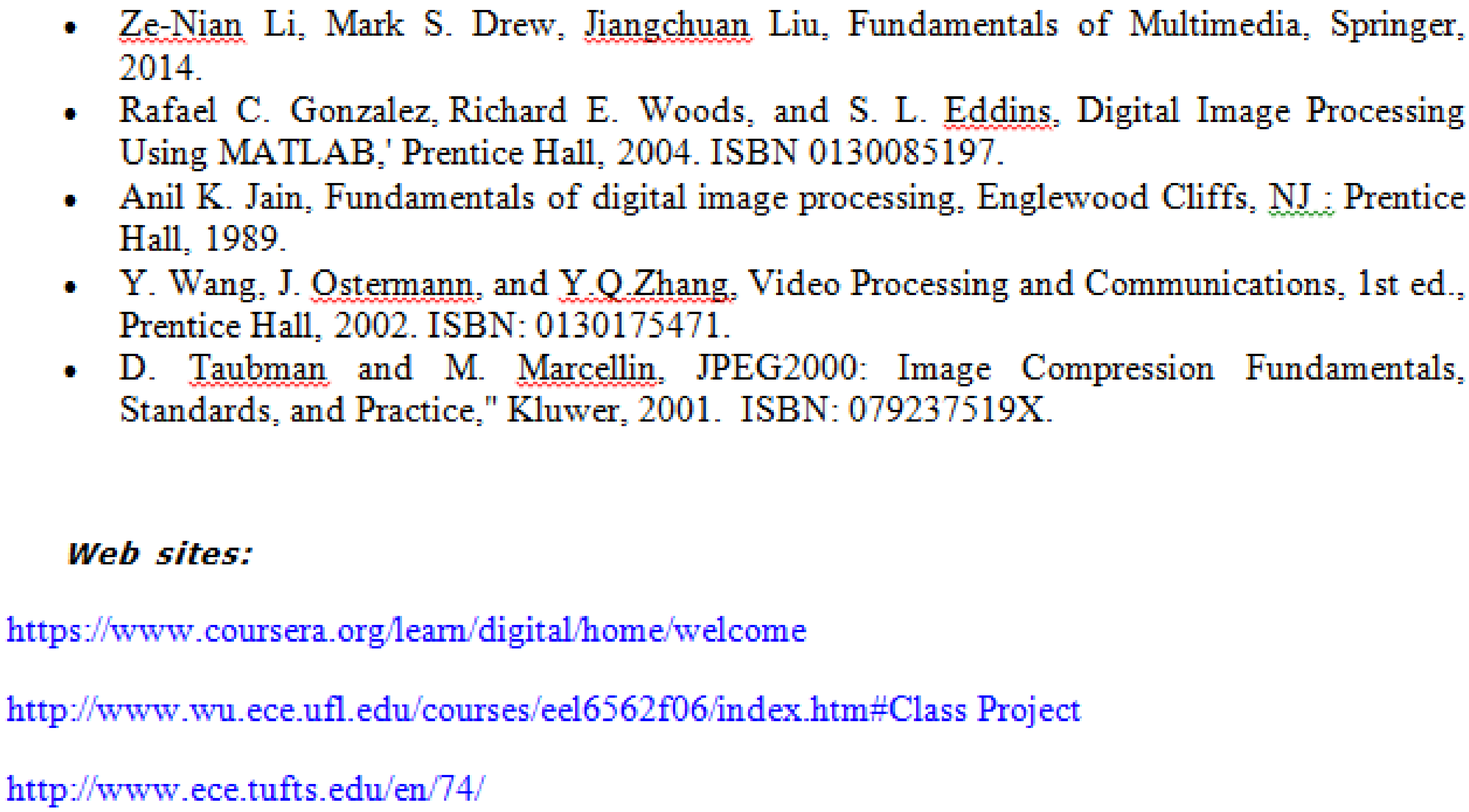


Week	Topics
1, 2, 3	Signal, image and video formation, image capturing, concept of multimedia, components of the multimedia system
4, 5, 6	Different artifacts and noise types that affect the images and videos, principles the image processing techniques, image and video enhancement,
7,8, 9, 10, 11	Midterm Exam Image and video restoration, segmentation, and classification techniques
12, 13	Different multimedia compression methods, performance evaluation of multimedia system
14	Mini-project discussion

Student Assessment

Assessment Method	Assessment Length	Schedule	Proportion
Written Examination	3h	On week 15	85 mark
Oral Assessment	--	--	--
Practical Examination	--	--	--
Semester work	5 hours (overall)	On week 3,5,6,9,12, 14	40 mark

Course Overview



- Ze-Nian Li, Mark S. Drew, Jiangchuan Liu, Fundamentals of Multimedia, Springer, 2014.
- Rafael C. Gonzalez, Richard E. Woods, and S. L. Eddins, Digital Image Processing Using MATLAB,' Prentice Hall, 2004. ISBN 0130085197.
- Anil K. Jain, Fundamentals of digital image processing, Englewood Cliffs, NJ : Prentice Hall, 1989.
- Y. Wang, J. Ostermann, and Y.Q.Zhang, Video Processing and Communications, 1st ed., Prentice Hall, 2002. ISBN: 0130175471.
- D. Taubman and M. Marcellin, JPEG2000: Image Compression Fundamentals, Standards, and Practice," Kluwer, 2001. ISBN: 079237519X.

Web sites:

https://www.coursera.org/learn/digital/home/welcome

http://www.wu.ece.ufl.edu/courses/eel6562f06/index.htm#Class Project

http://www.ece.tufts.edu/en/74/

What is Multimedia?

- The term "**multimedia**" have quite different definitions from different viewpoints.

- A consumer entertainment vendor, may think of multimedia as **interactive TV** with hundreds of digital channels, or a cable-TV-like service delivered over a high-speed Internet connection.

- Multimedia can be considered a **laptop** that has good sound capability.

- A computer science/engineering has a application-oriented view of what multimedia consists of: **applications that use multiple modalities to their advantage, including text, images, drawings, graphics, animation, video, sound (including speech), and, most likely, interactivity of some kind**.

What is Multimedia?

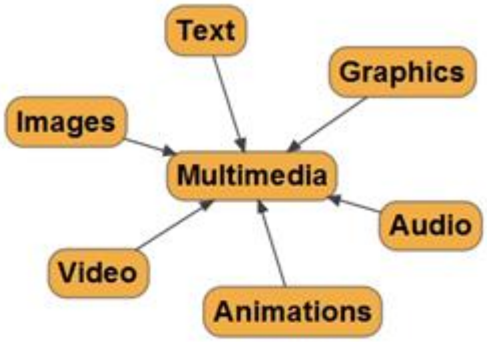
- This contrasts with media that use only fundamental computer displays such as text only or traditional forms of printed or hand-produced material.

- **Multimedia is part of some of the most interesting applications in computer science with *interaction*.**

- Graphics, visualization, computer vision, data compression, graph theory, networking, database systems—all have important contributions to make in multimedia at the present time.

Components of Multimedia

- Multiple forms of information, including text, audio, images, drawings, animation, video, and interactivity in multimedia have several applications, such as:
  - ✓ **Geographically based**, real-time augmented-reality, massively multiplayer online video games, making use of any portable device (smartphones, laptops, or tablets), which function as GPS-aware mobile game consoles.
  - ✓ Shapeshifting TV
  - ✓ Cooperative education environments
  - ✓ Searching large video and image databases for target visual objects using semantics of objects.
  - ✓ Compositing of artificial and natural video into hybrid scenes
  - ✓ Visual cues of video-conference participants



Multimedia

- **Multimedia processing and coding:**

Includes audio/image/video processing, compression algorithms, multimedia content analysis, content-based multimedia retrieval, and multimedia security.

- **Multimedia system support and networking:**

Includes network protocols, Internet and wireless networks, operating systems, servers and clients, and databases.

Multimedia

- **Multimedia tools, end systems, and applications:**

Includes hypermedia systems, user interfaces, multimodal interaction, and integration: Web-everywhere devices, multimedia education, including computer supported collaborative learning and design, and applications of virtual environments.

- **Multimedia production:**

It is easily involved an art director, graphic designer, production artist, producer, project manager, writer, user interface designer, sound designer, videographer, and 3D and 2D animators, as well as programmers.

Multimedia software tools

- ✓ Music sequencing and notation
- ✓ Digital audio
- ✓ Graphics and image editing
- ✓ Video editing
- ✓ Animation

Fundamentals of Related topics and concepts

Analog versus Digital

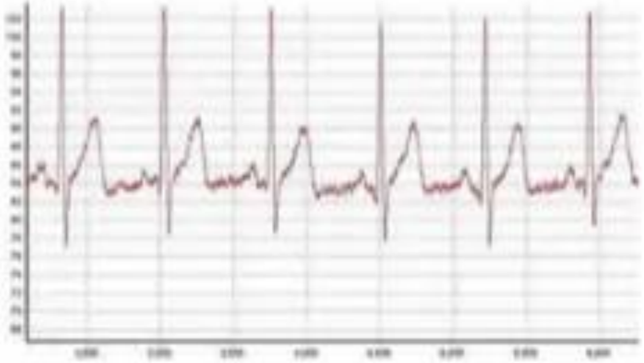
- We are surrounded by countless applications which make use of images and videos.

- End-to-end system has input and output in the analog worlds, but process and perform in the digital world.

- The concepts of sampling and quantization have a critical role.

Signal

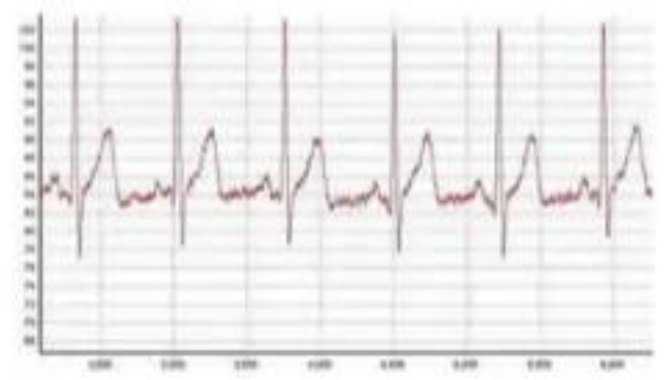
- A *function* containing *information* about the behavior or nature of some phenomenon of interest.
- In the physical world, any quantity exhibiting variation in time and/or space is potentially a signal.



Signal

- A signal can be defined as a function.
- One-dimensional, an x(t) for example function. t is the independent variable.
- Two-dimensional, x(t1,t2), where t1, t2 are the independent variables that contains information. It tells us something useful about the behavior or the nature of some phenomenon of interest.
- In the physical world, we can say that any quantity that changes with respect to time and/or with respect to space is potentially a signal.

Signal



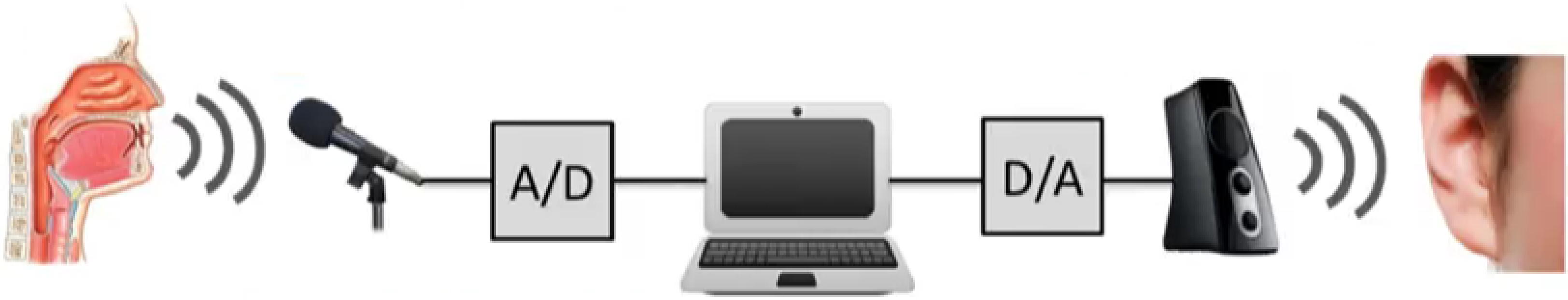
- An electrocardiogram signal is a signal example.

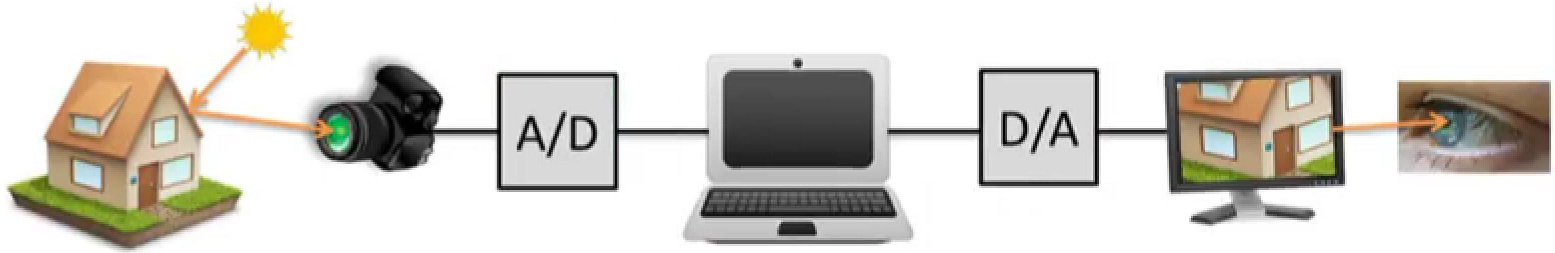
- Time on the horizontal axis and the amplitude of a signal on the vertical axis. So a signal like this shows the health of somebody's heart.

Signal

- Signals in systems play a very important role in many areas of science and technology.
- From communication to astronautics to circuit design, biomedical engineering and so on.
- Images and videos are the signals of interests we'll be dealing with.
- We have analog and digital signals. To put them in some perspective, consider the speech generation transmission processing perception system like the one depicted as follows.

Analog &Digital





Analog &Digital

- So speeches generated by the pressure provided by the lungs that become sound in the glottis in the larynx. This is then converted into vowels and consonants by the vocal track.
- Speech is an **acoustic signal** that is transmitted through sound waves utilizing the vibration of the air molecules.

Analog &Digital

- So such a signal reaches a microphone, which is a transducer. It converts one form of energy, acoustical energy, into another form of energy, electrical energy. So here I have an electrical signal.
- Both the signals are analog signals or continuous signals.
- This is an analog signal, because the independent variable t, as well as the amplitude x, which is a function of t are continuous numbers.

Analog &Digital

- Then, we have to process such a speech signal by a digital computer.

- Digital computers only understand zeroes and ones or understand digital signals. Therefore, it is essential to convert the analog signal into a digital one which is a function of this A to D converter box.

Analog &Digital

- In doing so, converting an analog to a digital signal, one has to take two steps.

- The first step is to sample the signal or discretize it in the time domain. This means that I look at the values of the signal at equally space points. So at value the time 0, the time capital T, the time 2T, 3T, 4T, minus T and so on.

Analog &Digital

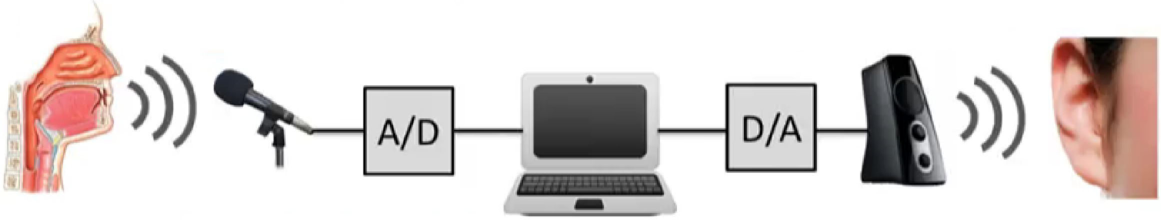
- So the signal values are obtained. Here at this time, at this time instance, this time instance and so on.
- Thus, the **Sampling process** is used to convert the signal into a **discrete time** signal.
- The second step is to discretize the amplitude of the signal to obtain only certain values available to represent x of t integer value. **[Quantization]**
- So let's assume that the values I have available are this one, this one, this one, this one. This means that, so let's call this 0, delta, two delta, three delta.

Quantization

- So this means that this value at 0 will be represented by this value. The second value will be represented by this value.
- Similarly, this one will be represented by this value. This will be represented by 0 and this one will also be represented by this value? So through this process that is called quantization.
- Thus, quantization output is a **discrete amplitude**.
- After both **sampling and quantization**, we end up with a signal that has both its independent variable as well as its amplitude being represented by these grid values >>> such a signal is called a **digital signal**.

Analog &Digital

- Now, this signal is an input to the computer which is going to process it and then at the other end, reverse order processes will be followed.
- In other words, the signal is turned from digital to analog through the converter. Then this analog signal reaches another transducer. This is also called an actuator that will turn the current into a vibration. Therefore, into another acoustic signal which is going to be perceived by the human auditory system. So here again, we have an electrical signal.

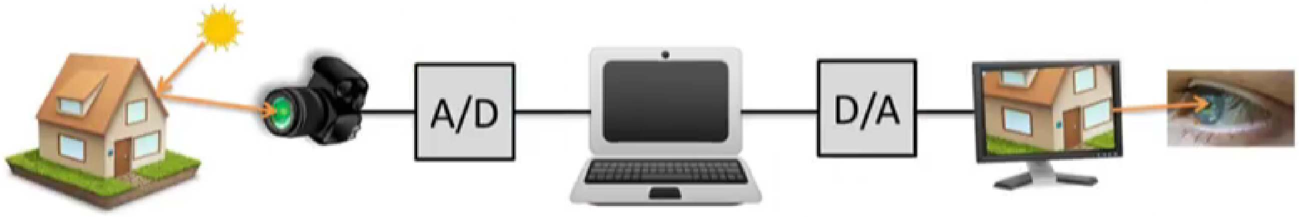


Analog &Digital

- And this is, and again, an acoustic signal. The objective of a digital signal processing is to focus onto the central part here.
- To focus on the techniques that would allow us to manipulate digital signals to perform certain tasks that depend on the application. So, we see here that we have two worlds, we have the Analog world here?
- So, analog, analog here, and the digital world or the digital domain here in the middle. That is of our primary interest.

Analog &Digital

- A similar situation to the previous one is depicted by this broad diagram that shows the generation, recording, processing, and sensing of an image.
- So, electromagnetic energy in the visible part of the spectrum is leaving the Sun reflected by the object, travels through the air and reaches a sensor.
- A sensor is a transduser which is an analog camera that through photo-chemistry converts light energy into chemical changes on the film.



Analog &Digital

- Using analog cameras, the brighter parts of the image lead to higher concentration of silver grains and this film, after it's processed, negative or black and white film, represents analog image.
- So our objective is to process the image by digital computer and therefore we have to convert the analog image to a digital one through this A to D block.
- In this particular case, we can use a densitometer that measures the concentration of the silver collide grains on the film or a scanner to end up with a digital image.

Analog &Digital

- After the digital image is processed by the computer, several techniques can allow us to process digital images.
- Thus, it is required to take the processed image to be seen by the human viewer. So the reverse part is followed.
- Whereby, the digital image is converted into an analog one and this signal will feed into a monitor, an analog monitor such as a CRT monitor that will convert the image into an electromagnetic wave that is going to reach the human visual system. The eye of the observer. Again, nowadays digital monitors or flat panels are prevalent. So an LCD here will be fed by the digital signal directly.

Analog &Digital

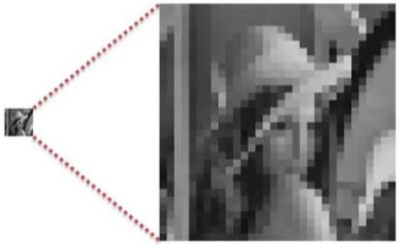
- A video adapter plays the role of the D to A converter.

- Images are represented by this A to D block, and actually could be resampling or re-quantization.

Sampling

- Consider a digital image of size **256 by 2**56 image, which means it has 256 rows and 256 columns and it's an eight bit image, which means that each element here which is called a pixel from picture element.

- Here, it is 2 to the eighth (one of 256 values).
- Down-sample image by a factor of eight in both the horizontal and the vertical directions.
- So out of **eight samples**, horizontally we will keep one sample, also the same is done in the vertical direction, i.e. eight by eight block.



Down- sampling

- Thus, we only keep this value and throw away the rest, which is 63.

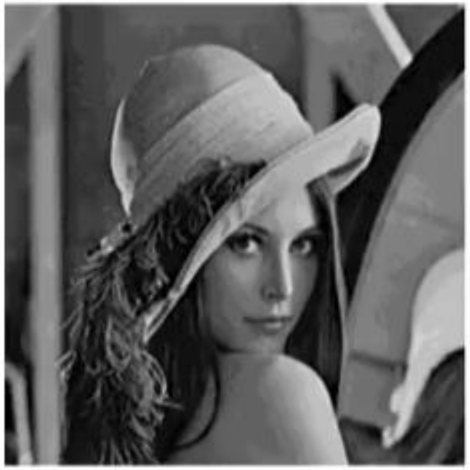
- By down sampling, we will end up with a **32 by 32 image**. This tiny image!!

- If we bring it back to the 256 by 256 dimension for visualization purposes, then that's how this image will look like. So, due to this down-sampling, blocking artifacts called jagged edges, appear. They're not straight edges anymore, but they're jagged.

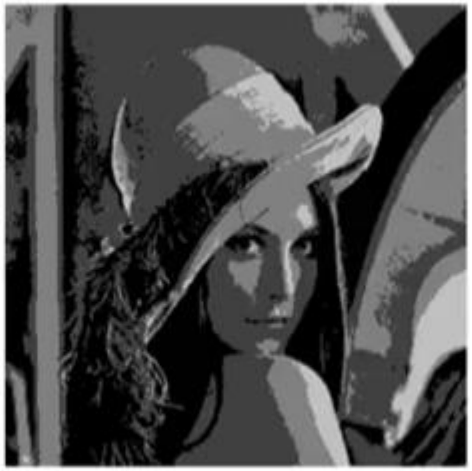
- However, up-sample the image will result in zero-order hold.

Quantization

8 bits per pixel



4 bits per pixel



2 bits per pixel

Re-quantization

- Similarly, the idea of de-quantization is depicted.

- **Contouring effects** appear, where artificial contours or boundaries appear on the original image.

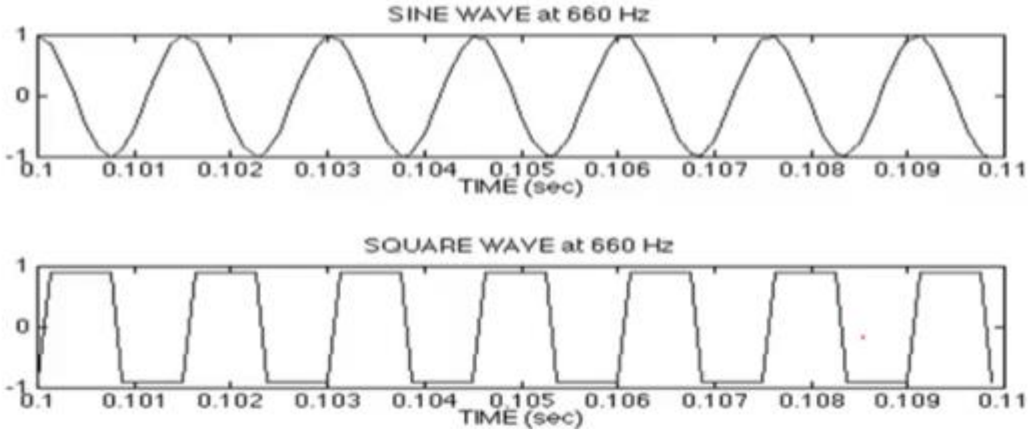
- Quantization errors.

Images and Videos

Signals are depending on the number of the independent variables, thus can be: 1D, 2D, 3D, and multidimensional signals:

- 1D: speech, audio, biomedical signals, etc. (function of time, which are continues or discreet having one independent variable)
- 2D: text, images, etc.
- 3D: video, volume, etc.
- MD: video of a volume

1D Signals: Tones



- Tones are examples of one-dimensional signals. A sine wave that only has one frequency present in it, at 660 hertz. That's why it's called a pure tone.
- At the bottom, a square wave appear. It also has the frequency at 660 hertz.
- However, it has additional frequencies, the so-called harmonic.
- It should be clear that the bottom square wave is richer in the sense of having additional frequencies down the proton the sinus wave.

2D Signals: Text & gray image

- Text and grayscale images are certainly two dimensional signals, where they're independent variables to present space.

- Color, multispectral images have two special coordinates, and then one can look at the amplitude in two different ways.

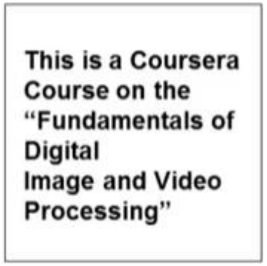
Images

- There are representative examples of images with different intensity **resolutions**.

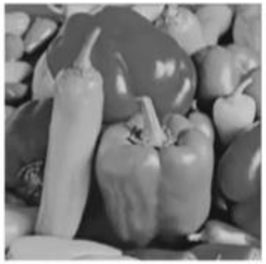
- For example, binary images such as a text image, through color images, when each pixel assumes one of 60 million colors, and pseudo color images.

- When we decide the number of colors to be used for visualization purposes for a color but also for a black and white image, there's pseudo coloring problem.

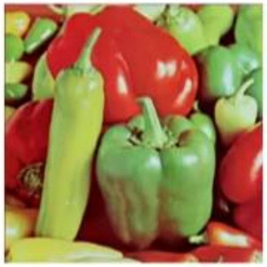
2D: Images


This is a Coursera
Course on the
“Fundamentals of
Digital
Image and Video
Processing”

1 bit per pixel



8 bits per pixel

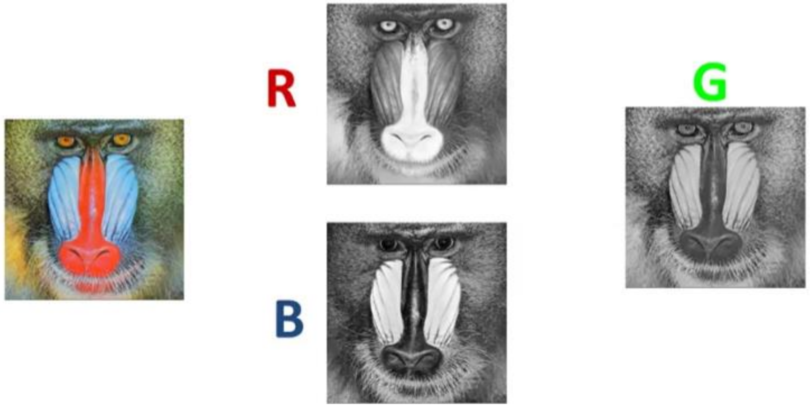


24 bits per pixel

- On the left, a **binary image** is represented having two bits to represent the different colors, we have a black and a white value here.

- In the middle, an **eight bits per pixel image**, while on the right is a **24 bit per pixel image**. This is a true color image. We have two to, to the 24 different color values.

RGB Images

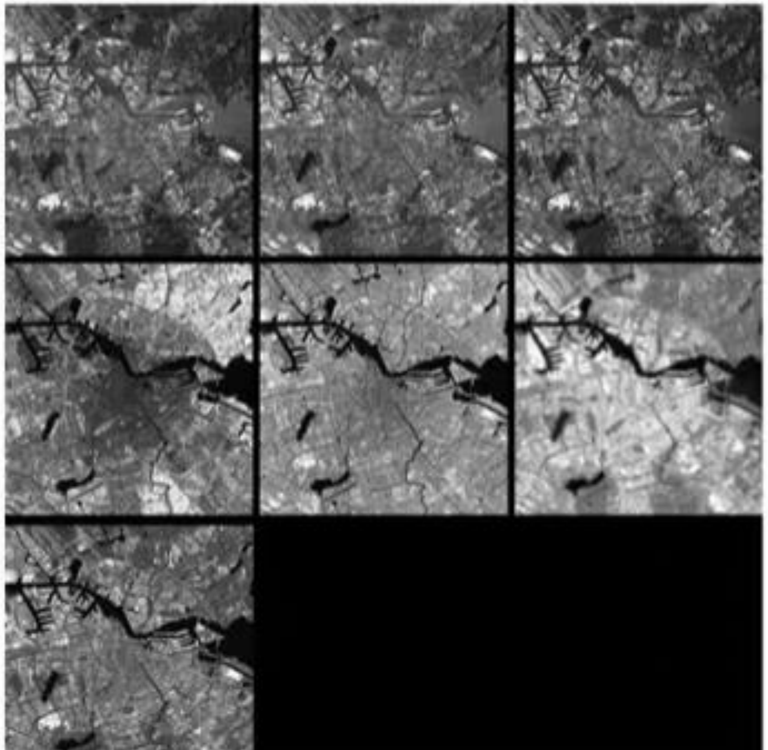


- There are different ways to represent the color image, one of which is shown here, in terms of three different channels, the red, green and blue channels.
- This is an RGB decomposition of the image. So, each of these channels is a black and white image. We use eight bits to represent it, so each of the channels here has two, 256 different values and eight bits per channel times three, therefore 24 bits to represent the color image.
- So, if we look at this you notice the nose of the mandrill here is quite red. And therefore, we see that the pixel values in the red channel are quite high. Why it is represented by height? Values closer to 56, while 0-55 so close to 55 while the darker values are closer to zero.
- So, the red nose has high values in the red channel and pretty small values in the other three channels.
- On the other hand if we look at the cheeks of the mandrill this is the variation of blue not exactly the blue color, so we see high values of red, in the blue channel and there are the smaller values, well definitely smaller than the red and semi somewhere in the middle in the green channel. Actually this particular value apparently is a combination of green and blue.

Landsat Images

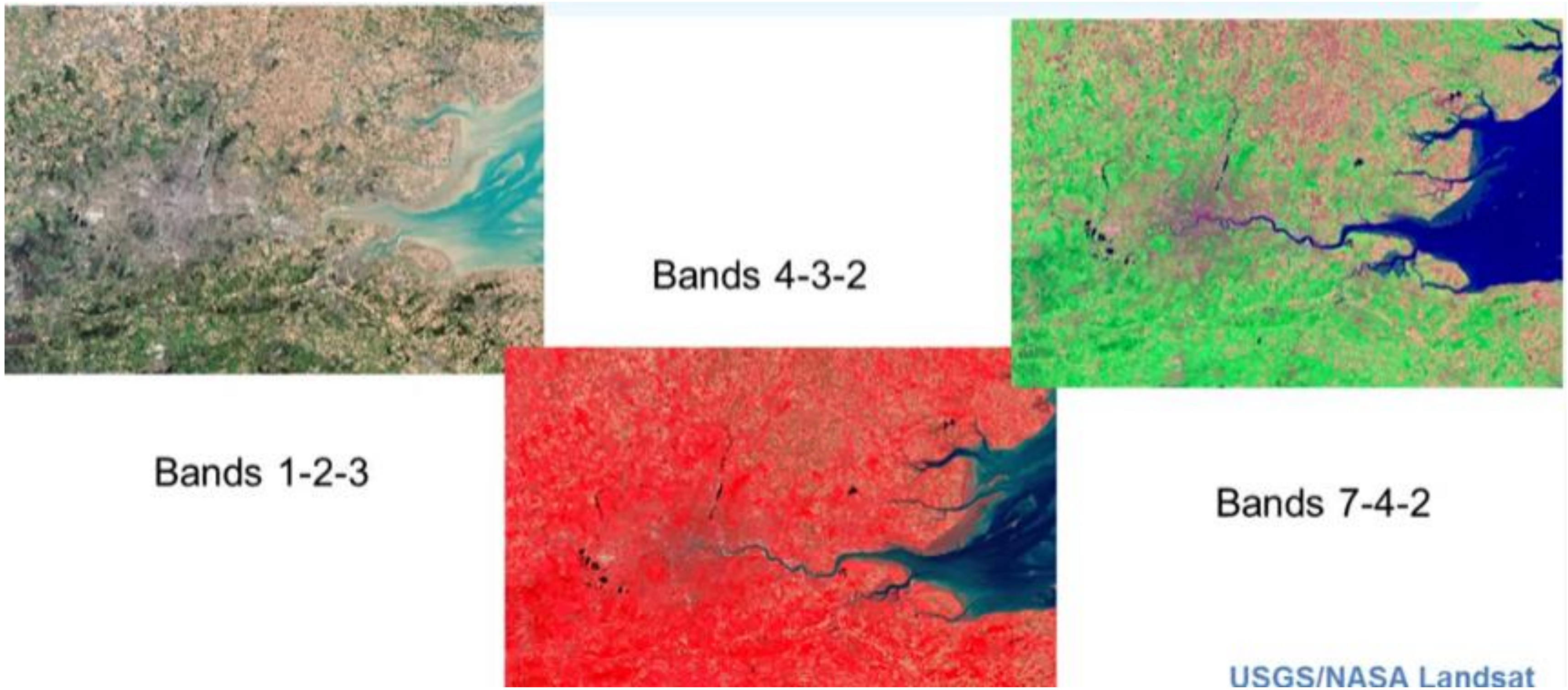
- The Landsat program is responsible for the acquisition of satellite imagery (earth).

- As shown, Landsat images with Multi-spectral views.



Hyperspectral images

- The hyperspectral images are the ones that have many more bonds, up to 200 and 300 bonds, but the main characteristic of the bonds are much closer spaced.



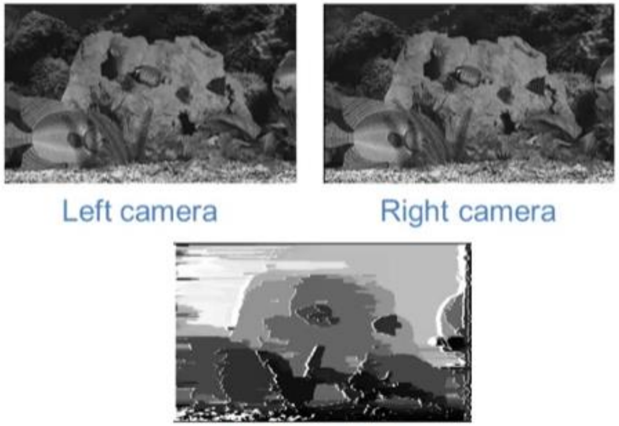
USGS/NASA Landsat

Stereo Images and Disparity

- The natural world is three dimensional while the images are two dimensional. They present the projection of a three D world on to the two D plane.

- Stereo images following the human perception that is variable to perceived depth.

- Using two cameras at some reasonable distance, the so-called base line, then the two images or videos recorded by the camera emulate our human visual system.
- so, we need to cameras that will capture the same scene, and therefore they will emulate the human visual system.
- In the previous figure, the left images shows what the left camera sees, and on the right, what the right camera sees.



Left camera

Right camera

Disparity Map

Stereo Images and Disparity

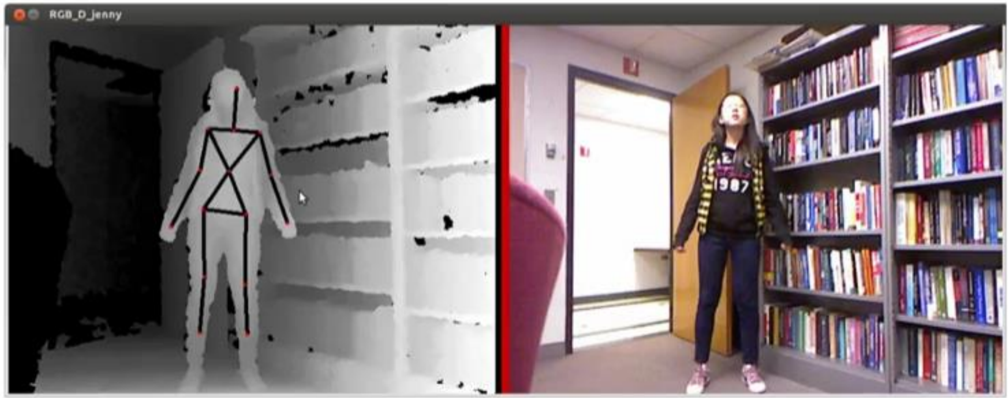
- The difference between these two images is the so-called **disparity map**, that represents how each and every pixel moved from one image, and going from one image to the next.

- The disparity map relates to the depth in the image. So the two channels infuse the property will give us the depth perception, and the fusion could be done through a red and blue channel and therefore we can use the color coded glasses or we could use glasses with different polarization and so on.

Kinect Images

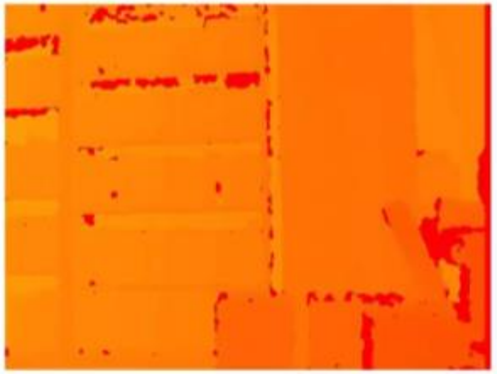
- Kinect camera can be used to obtain depth information. It projects a known pattern onto the scene and infers depth from the information from the pattern. So the kinect combines **structural light** with two computer vision techniques depth from **focus** and **depth** from stereo.

- The kinect is intended to use with the Xbox and therefore it is of interest there to compute the depth map but also infer the body position. So you see here the skeleton of the person on, on the left image, right. So here is the person, this is the depth map, and also the kinect provides a visible image shown here. So you see the visible and the corresponding depth image that showed where the person stands and where the door is and so on.

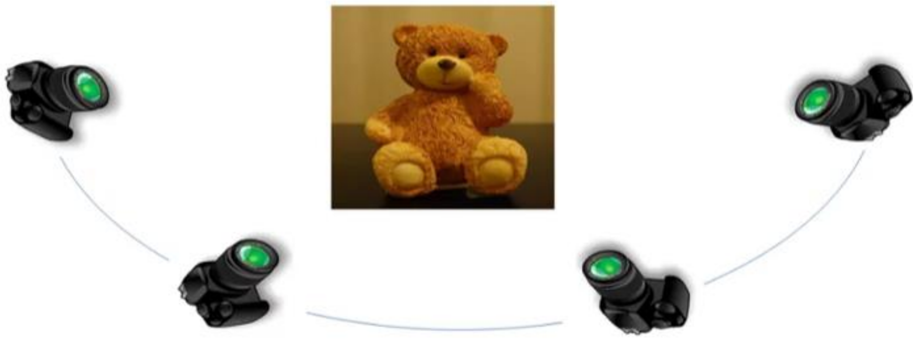


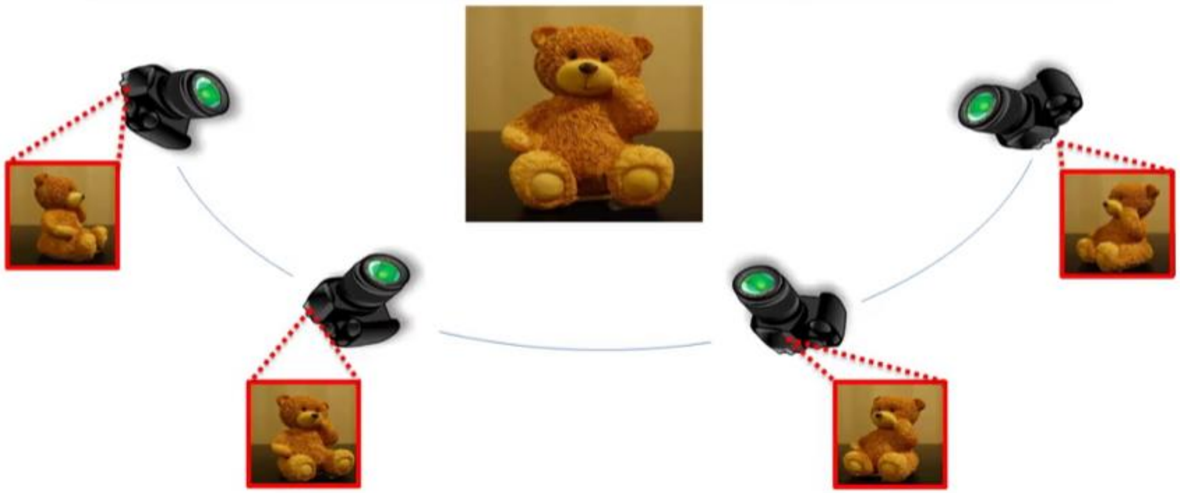
3D: Videos

- Video is a 3D signal. It has two spatial and one temporal coordinates, while a 3D volume has three spatial coordinate x, y, z.
- Video can show both the visible image as well as the depth image acquired by the kinect camera.
- Instead of using two cameras as in the stereo case, we use many cameras on a specific right. So, the image of this particular object is viewed from many different angles.
- As an example of a four dimensional signal >> is looking at the objects volume.



Multi-Camera Imaging





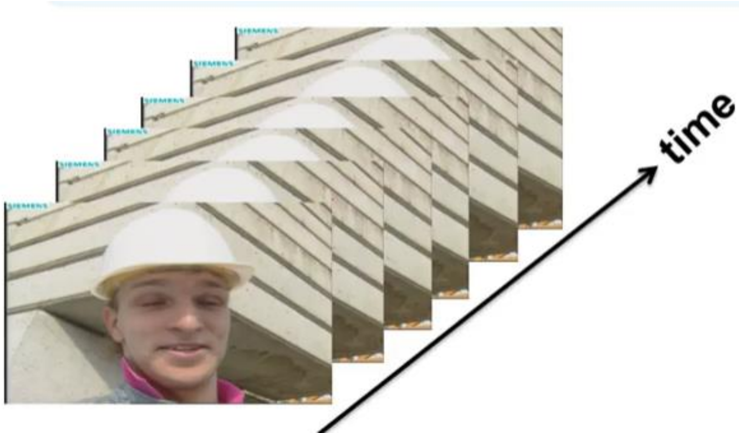
4D Representation

- Typically, x, y, z signal that changes over time, so time is the fourth independent variable.

- Some of the tools that we use to describe signals carry over from 1D to 2D to MD is the straight forward extension one just adds one more variable, and everything remains the same in some sense.

Videos

- A video consists of **individual frames**, and one could argue that a video is nothing else than a **collection of images**.
- Therefore, if we have an outgoing that is effective in processing an **image**, a **still frame**, as it's called. Also, we could apply the same algorithm to frame after frame.
- However, what is special about video is that these frames are highly correlated and therefore, we can gain if in processing such frames. Thus, this **correlation** should be taken into account. Of course these frames, if they're displayed at some frame rate, 30 frames per second, for example, one can perceive the actual motion indices.



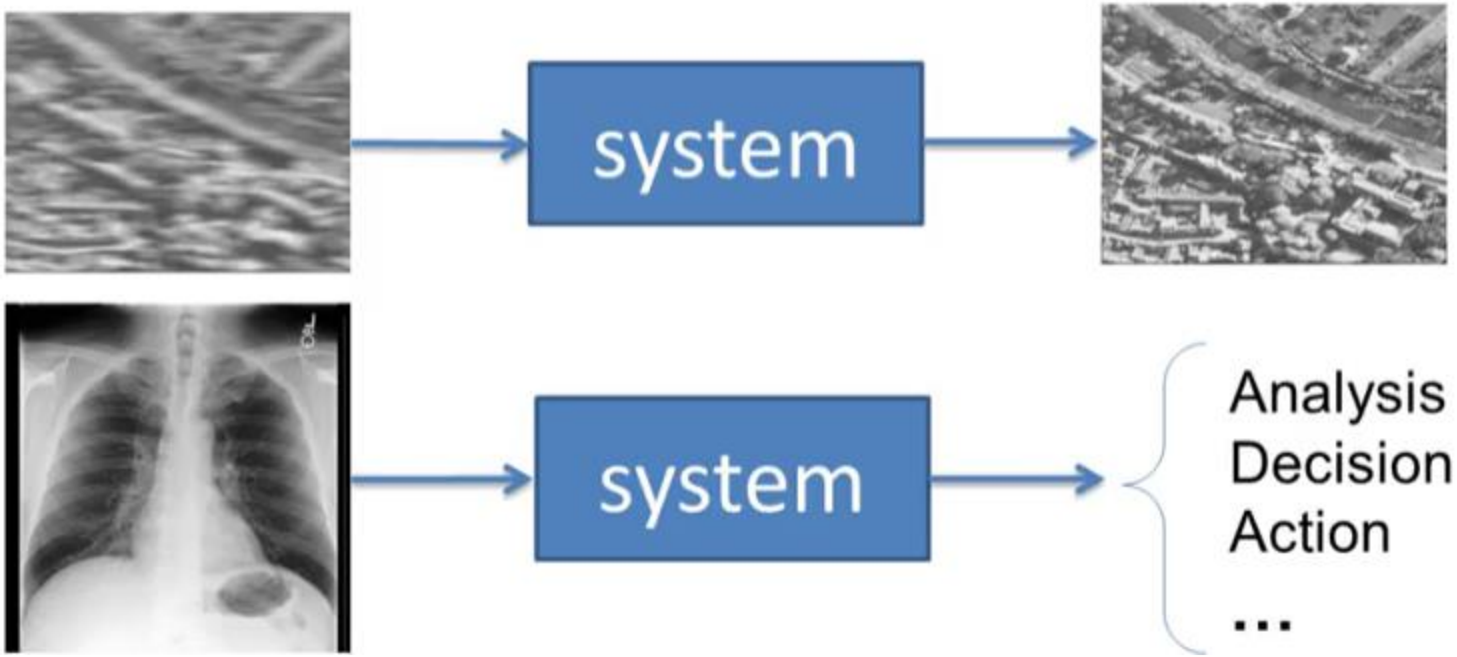
Processing system

- The narrow definition of a **processing system** is one that accepts an image at the input and generates another image at the output.
- It is a system which for an image at the input might generate an image at the output or a decision based on the analysis of the image, which might result in an action such as to operate for example in a patient if there is a medical image.
- Or the extraction or segmentation of an object from an image based on its color or it's motion and maybe even its classification.

Processing System

- Images and videos are clearly the focus of the course, where the images are two dimensional signals but can also be three dimensional are in the multispectral, hyperspectral images while video is a three dimensional signal. So we are a dimension.
- Finally, **processing** means the manipulation of the values of an image or a video by computer.
- The result of processing might be the removal of blur, as is the case here you, you see, the input to the system is an aerial photograph that is blurred with the motion between the camera and the scene and the output is a sharpened, a restored image.

Image and Video Processing



system

system

Analysis
Decision
Action
...

- Now if an image is input, an image is output this has been the kind of narrow definition of processing, or so but if we do as filtering, and the broader meaning of processing that we adopt here is that an image or video can be input to a system and we're interested in **extracting important features** from such an image or we're interested in making decisions based on the image.
- The best example here shows this is a chest x-ray and the input and based on the analysis that would be performed whether there's a malignant tumor or not, for example, certain decisions and actions will be the output of the system.

Image and Video Processing

- As time goes on, what we see is that the boundary is between tradition and separate areas become **fuzzy** or in other words there's overlap between these traditionally separate areas. So, when it comes to signal processing and more specifically between with the video processing that we'll be dealing with in this class. There is overlap between the, this field and the fields of communication, computer vision, machine learning and optimization.

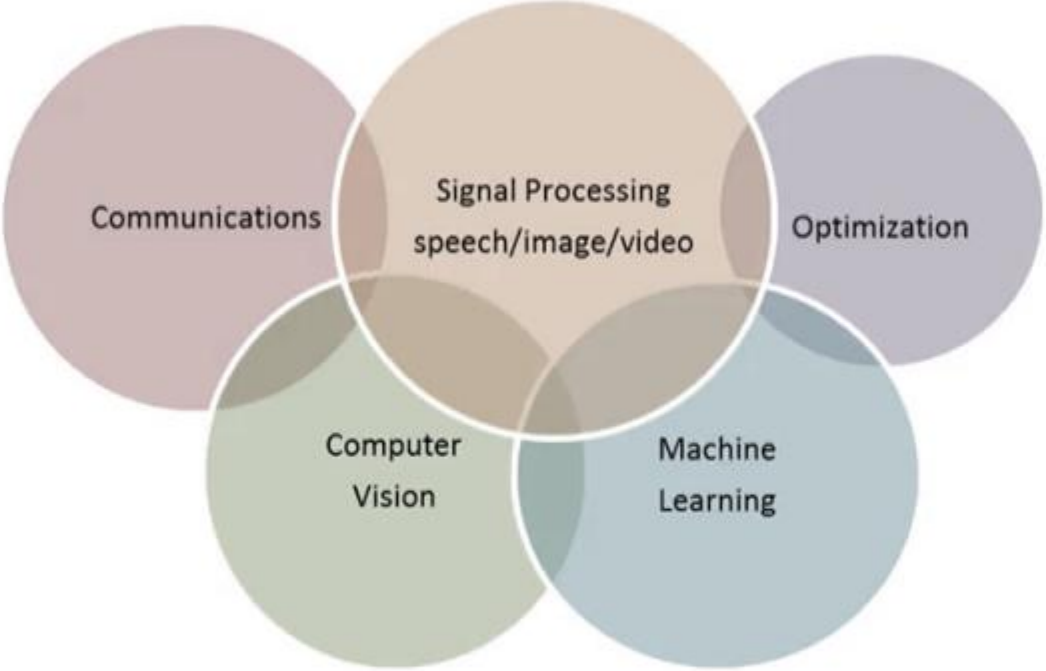


Image Processing

- Modern digital technology has made it possible to manipulate multi-dimensional signals with systems that range from simple digital circuits to advanced parallel computers.
- The goal of this manipulation can be divided into three categories:

❑Image Processing *image in → image out*

❑Image Analysis *image in → measurements out*

❑Image Understanding *image in → high-level description out*

Image Processing

- An image defined in the “real world” is considered to be a function of two real variables, for example, $a(x,y)$ with a as the amplitude (e.g. brightness) of the image at the *real* coordinate position (x,y).
- An image may be considered to contain sub-images sometimes referred to as *regions–of–interest*, *ROIs*, or simply *regions*.
- This concept reflects the fact that images frequently contain collections of objects each of which can be the basis for a region.
- In a sophisticated image processing system it should be possible to apply specific image processing operations to selected regions. Thus one part of an image (region) might be processed to suppress motion blur while another part might be processed to improve color rendition.

Image Processing

- The amplitudes of a given image will almost always be either real numbers or integer numbers. The latter is usually a result of a quantization process that converts a continuous range (say, between 0 and 100%) to a discrete number of levels.
- In certain image-forming processes, however, the signal may involve photon counting which implies that the amplitude would be inherently quantized.
- In other image forming procedures, such as magnetic resonance imaging, the direct physical measurement yields a complex number in the form of a real magnitude and a real phase.

Digital Image Definitions

- A digital image *a*[*m*,*n*] described in a 2D discrete space is derived from an analog image *a*(*x*,*y*) in a 2D continuous space through a *sampling* process that is frequently referred to as digitization.

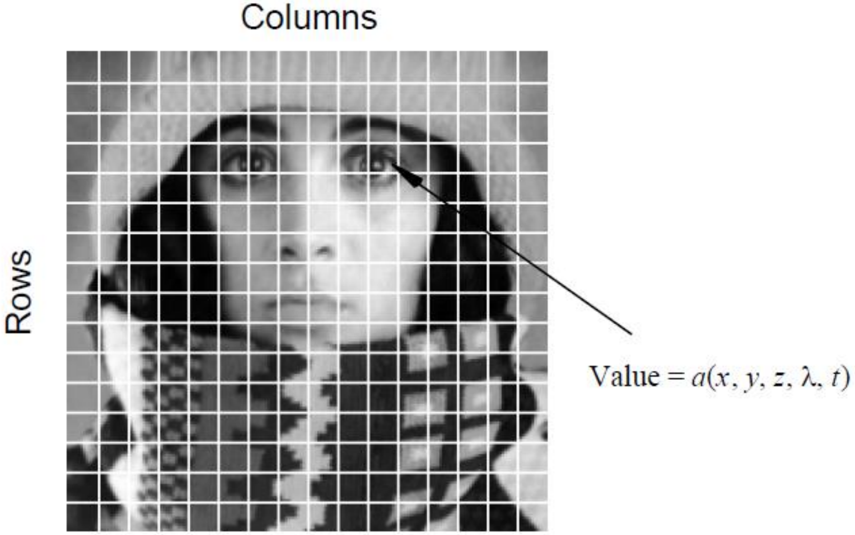
- The 2D continuous image *a*(*x*,*y*) is divided into *N rows* and *M columns*. The intersection of a row and a column is termed a *pixel*.

- The value assigned to the integer coordinates [*m*,*n*] with {*m*=0,1,2,…,*M*–1} and {*n*=0,1,2,…,*N*–1} is *a*[*m*,*n*].

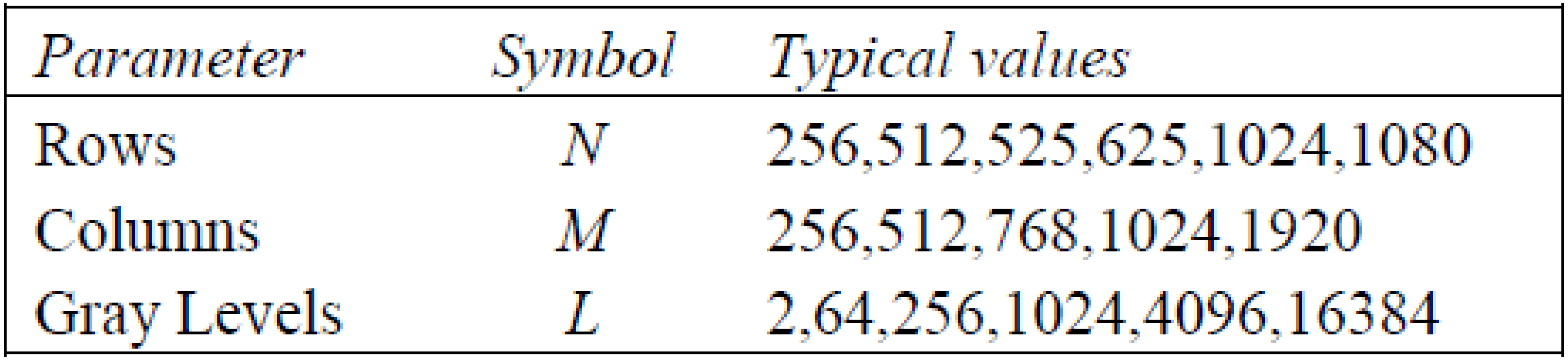
- In fact, in most cases *a*(*x*,*y*) – which we might consider to be the physical signal that impinges on the face of a 2D sensor – is actually a function of many variables including depth (*z*), color (λ), and time (*t*).

Digital Image Definitions

- The image shown in the Figure has been divided into *N* = 16 rows and *M* = 16 columns.
- The value assigned to every pixel is the average brightness in the pixel rounded to the nearest integer value.
- The process of representing the amplitude of the 2D signal at a given coordinate as an integer value with *L* different gray levels is usually referred to as amplitude quantization or simply *quantization*.



Common Values

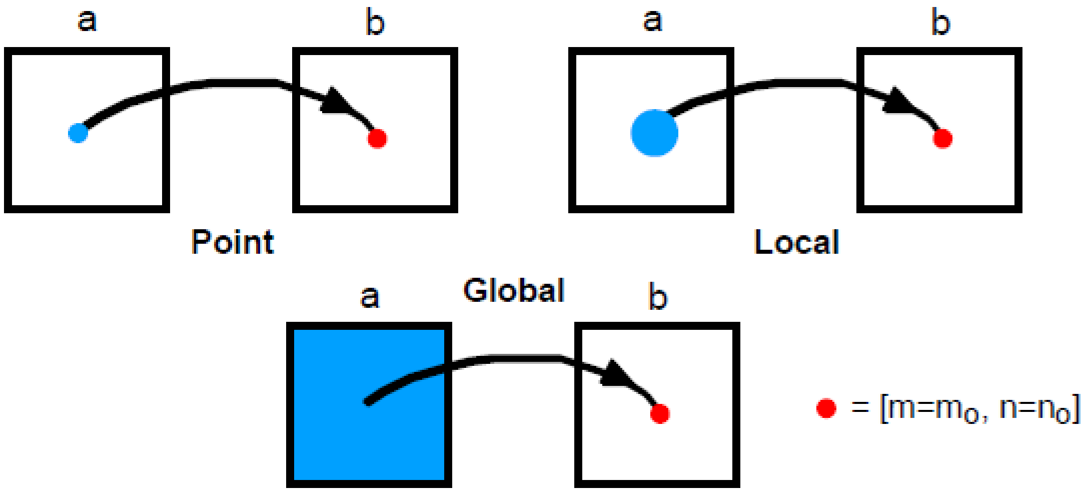


Parameter	*Symbol*	*Typical values*
Rows	*N*	256,512,525,625,1024,1080
Columns	*M*	256,512,768,1024,1920
Gray Levels	*L*	2,64,256,1024,4096,16384

- There are standard values for the various parameters encountered in digital image processing. These values can be caused by video standards to keep digital circuitry simple, where *M*=*N*=2^*K* where {*K* = 8,9,10,11,12}.

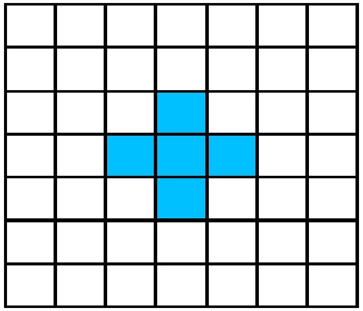
Characteristics of Image Operations

Operation	Characterization	Generic Complexity/Pixel
• *Point*	– the output value at a specific coordinate is dependent only on the input value at that same coordinate.	*constant*
• *Local*	– the output value at a specific coordinate is dependent on the input values in the *neighborhood* of that same coordinate.	P^2
• *Global*	– the output value at a specific coordinate is dependent on all the values in the input image.	N^2

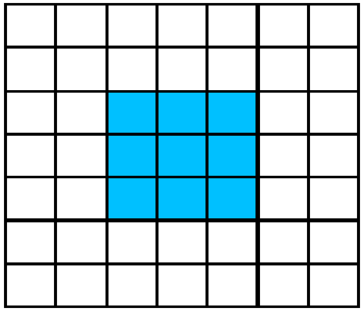


Types of neighborhoods

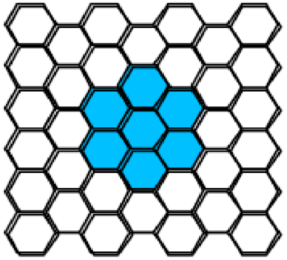
- Neighborhood operations play a key role in modern digital image processing. It is therefore important to understand how images can be sampled and how that relates to the various neighborhoods that can be used to process an image.
    - □ **Rectangular sampling** – In most cases, images are sampled by laying a rectangular grid over an image.
    - □ **Hexagonal sampling**
- Local operations produce an output pixel value *b*[*m*=*mo*,*n*=*no*] based upon the pixel values in the *neighborhood* of *a*[*m*=*mo*,*n*=*no*].
- Some of the most common neighborhoods are the 4-connected neighborhood and the 8-connected neighborhood in the case of rectangular sampling and the 6-connected neighborhood in the case of hexagonal sampling .



Rectangular sampling
4-connected

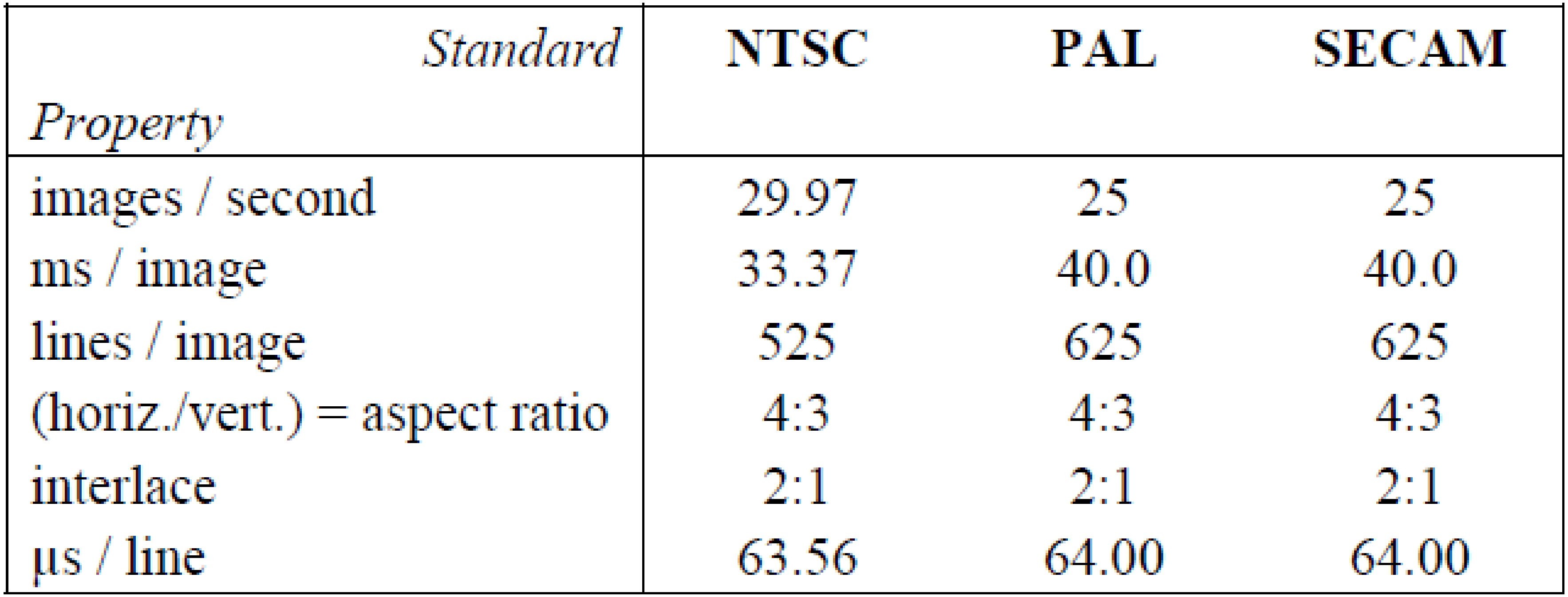


Rectangular sampling
8-connected



Hexagonal sampling
6-connected

Video Parameters



Standard	**NTSC**	**PAL**	**SECAM**
Property			
images / second	29.97	25	25
ms / image	33.37	40.0	40.0
lines / image	525	625	625
(horiz./vert.) = aspect ratio	4:3	4:3	4:3
interlace	2:1	2:1	2:1
µs / line	63.56	64.00	64.00

References

- https://www.coursera.org/learn/digital

- Ze-Nian Li, Mark S. Drew, Jiangchuan Liu, Fundamentals of Multimedia, Springer, Second Edition, 2014.